Article

Evaluating Design Guidelines for Intuitive, Therefore Sustainable, Virtual Reality Authoring Tools

Iolanda L. Chamusca [1,*], **Yiyu Cai** [2], **Pedro M. C. Silva** [3], **Cristiano V. Ferreira** [4], **Thiago B. Murari** [5,6], **Antonio L. Apolinario, Jr.** [7] **and Ingrid Winkler** [5,6,*]

1   Ford Motor Company, Camaçari 42831-710, Brazil
2   School of Mechanical & Aerospace Engineering, College of Engineering, Nanyang Technological University, Singapore 639798, Singapore; myycai@ntu.edu.sg
3   NVIDIA Brazil & Latin America, Rua George Ohm, 206, São Paulo 04576-020, Brazil; pcruzesilva@nvidia.com
4   Technological Center of Joinville, Federal University of Santa Catarina, Rua Dona Francisca, 8300, Zona Industrial Norte, Joinville 89219-600, Brazil; cristiano.v.ferreira@ufsc.br
5   Department of Industrial Management and Technology, SENAI CIMATEC University Center, Salvador 41650-010, Brazil; thiago.murari@fieb.org.br
6   Institute for Science, Innovation and Technology in Industry 4.0/INCITE INDUSTRIA 4.0, Salvador 41650-010, Brazil
7   Department of Computer Science, Federal University of Bahia, Salvador 40170-110, Brazil; antonio.apolinario@ufba.br
*   Correspondence: ichamusc@ford.com (I.L.C.); ingrid.winkler@doc.senaicimatec.edu.br (I.W.)

Abstract: Virtual reality software might be challenging to utilize for beginners and unskilled professionals who do not have a programming or 3D modeling background. Concurrently, there is a knowledge gap in software project design for intuitive virtual reality authoring tools, which were supposed to be easier to use. These tools are frequently insufficient due to a lack of support and standard operating procedures. This study evaluates the validity of fourteen design guidelines for the development of intuitive virtual reality authoring tools. Adopting the Design Science Research approach, a previous study completed the first steps of the protocol by identifying problems, defining solution objectives, and developing and demonstrating the design guidelines. This work evaluates their application in a experiment. A group of engineering students with no prior experience in creating virtual worlds were tasked with examining the design guidelines while using the NVIDIA Omniverse Enterprise as an exemplary use case by following the tutorials in the platform. Previously, the students answered a Likert-scale questionnaire and a focus group interview, with eighteen questions about how they perceived these guidelines. An average score value for the tool was estimated through the questionnaire answers and a correlation analysis and, using the Pearson Correlation Coefficient (PCC), it was confirmed that most guidelines scores behaved as expected and were ranked according to the use-case functionality. The participants understood the guidelines' definition and could decide if they agreed or disagreed with their presence during the experiment. We evaluated that, in accordance with the Design Science Research, the proposed artifact is useful, i.e., the design guidelines for virtual reality authoring tools perform what they are designed to do and are operationally reliable in accomplishing their goals.

Keywords: virtual reality; authoring tools; NVIDIA Omniverse; intuitiveness; user-centered design; human-computer interaction; design guidelines

1. Introduction

Virtual reality (VR) has contributed to the achievement of the United Nations' Sustainable Development Goals (SDGs) in several ways. Notable examples [1] include SDG 4's inclusive and equitable education for all, with immersive learning experiences through virtual classrooms, simulations, and educational games, enhancing accessibility and engagement for diverse learners and promoting global collaboration and cultural exchange.

VR facilitates global collaboration, connecting students and educators worldwide, breaking down geographical barriers, and fostering cultural exchange. In addressing SDG 3 (Good Health and Well-being), virtual reality significantly advances healthcare. VR and AR technologies enable healthcare professionals to practice surgeries and procedures without risk, improving skills and patient outcomes. For SDG 8 (Decent Work and Economic Growth), virtual reality creates new work and entrepreneurship opportunities. Virtual marketplaces enable the trade of virtual goods and services, contributing to virtual economies and income opportunities, particularly in developing countries. Remote work and collaboration are facilitated, exemplified by platforms like Virbela, reducing geographical barriers and promoting economic inclusion. Regarding SDG 13 (Climate Action) and SDG 15 (Life on Land), virtual reality raises environmental awareness through virtual simulations that demonstrate human impact on the environment and the consequences of climate change, encouraging sustainable behaviors. Collaborative environmental initiatives in virtual reality, like the Virtual Reality Climate Research Hub, unite experts and communities in climate action and biodiversity conservation. SDG 5 (Gender Equality) and SDG 10 (Reduced Inequality) highlight virtual reality's role in promoting equality by offering an inclusive platform for diverse voices. Virtual spaces enable participation from varied backgrounds, fostering cultural exchange and appreciation. SDG 16 (Peace, Justice, and Strong Institutions) also benefits from virtual reality's capacity for international dialogue and conflict resolution, transcending physical boundaries and enhancing the global community sense. Platforms like United Nations Virtual Reality (UNVR) facilitate diplomatic efforts and governance transparency, contributing to stronger institutions. [1].

In the context of digital transformation, VR plays a pivotal role in achieving SDG 12 (Responsible Consumption and Production) by influencing consumer behavior towards virtual consumption. This shift reduces the environmental footprint traditionally associated with physical production and consumption. Virtual goods and services, designed with sustainability at their core, encourage eco-friendly choices, contributing to decreased resource use and waste, in line with sustainable development principles [1]. Furthermore, VR supports sustainable manufacturing processes and aids organizations in efficiently managing and allocating resources, with an emphasis on carbon neutrality and green development [2]. Additionally, SDG 9 emphasizes the importance of resilient infrastructure and innovation, where the immersive experience provided by VR and the metaverse stands as a significant innovation. Investments in metaverse infrastructure, such as high-speed internet and advanced computing, enable technological innovation. This fosters economic growth, encourages sustainable industrial practices, and aids in building resilient infrastructure [1], showcasing VR's integral role in advancing these global goals.

This cross-sectoral deployment of VR technology illustrates its significant potential as a tool in promoting sustainable development and addressing global challenges.

However, the development of virtual reality experiences remains an expensive and time-consuming process, relying on robust game engines that require the expertise of skilled professionals [3–5]. Virtual reality software architecture is complicated and involves a broad spectrum of assets, involving atypical input and output components such as head-mounted displays (HMDs), and tracking systems, among others [6,7]. Moreover, immersive technology is complicated and professionals need to have a wide range of skills, including a lot of technical knowledge in programming languages and/or 3D modeling [5,8–10]. So, creating interactive scenes in virtual reality is still complex and uncomfortable for people who have never done it before [9,10]. In order to foster a more sustainable world, it is imperative not only to expand access to virtual reality experiences but also to empower more individuals with the ability to create them. Developing sustainable virtual reality technology is important for attaining a sustainable future [11].

Authoring tools are an alternative to the lengthy learning curve, as they aim to facilitate the creation of content with minimal iterations. The term *authoring tool* refers to software structures that include only the most relevant tools and features for content creation while enhancing and speeding up product maintenance [12,13]. Unlike high-fidelity prototypes

that need substantial understanding of coding, these technologies can be employed for low-fidelity writing, which requires less coding experience [9]. Virtual reality experiences can presently be created using a variety of authoring tools, many of which are free-source programs [14]. But these tools usually lack documentation and tutorials in addition to functionality, which makes them unsuitable for supporting the complete development cycle [9,15].

Professionals of all skill levels would benefit from mature and mainstream authoring tools that are intuitive, helping them reach their virtual reality goals more quickly. In addition, the fast advancement of immersive technology may promote principles like the metaverse, which allows users to create digital works backed by the metaverse engine and live a seamless virtual life, especially with the help of human-computer interaction (HCI) and extended reality (XR) [16]. Similar to authoring tools, integrated virtual world platforms (IVWPs), like Roblox, Minecraft, and Fortnite Creative, are used to create games through graphical symbols and objectives instead of code and have a simpler interface, enabling users to create virtual worlds for the metaverse with less support, money, expertise, and skills [17].

On the other hand, software or platforms in the form of authoring tools are very hard to develop because they aim to give creators creative freedom while standardizing underlying technologies, making everything as interconnected as possible, and minimizing the need for creators to be trained or know how to program [17]. In the end, every feature becomes a priority.

This issue has previously been addressed, and design guidelines have been compiled to assist software developers in defining authoring tool projects [11]. These guidelines were intended to assist such developers in selecting and formulating the *features* and *requirements* that the authoring tools would need to satisfy to be deemed intuitive [18]. Additionally, they served as a means for virtual reality creators to assess the level of intuitiveness of tools that had already been developed. Figure 1 illustrates the information flow when using design guidelines in these two scenarios.
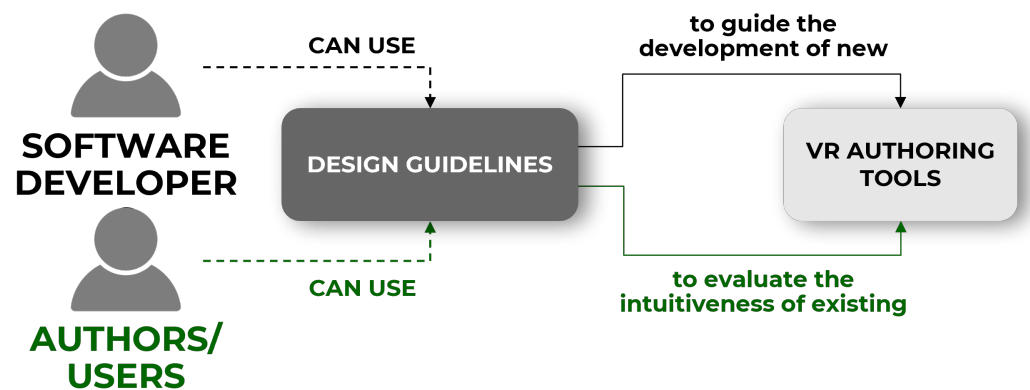

Figure 1. The design guidelines' artifact may be used at two stages of the life cycle of a VR authoring tool (adapted from Chamusca et al. [11]).

Chamusca et al. [11] developed and demonstrated the design guidelines as an artifact, but they have not yet been assessed. According to the Design Science Research (DSR) paradigm, it is important to collect evidence that a proposed artifact is useful. This means showing that the proposed artifact works and does what it is supposed to do, i.e., that it is operationally reliable in achieving its goals [19].

The most straightforward and effective method for creating the metaverse and enabling the transfer of data, virtual products, and money between various virtual worlds is still up for discussion. Yet, these design principles make it possible to develop virtual reality authoring tools that are easier to use. As the global economy shifts to virtual worlds, virtual world engines will become standard elements of the metaverse [17]. Our study ad-

dresses the challenging task of making VR software in general more accessible to beginners and non-experts, addressing a gap in the software design for intuitive VR tools. The key contributions of our study can be summarized as follows:

1. Providing important insights into the intuitiveness of the VR authoring tool and highlighting the significance of user-centered design in VR software development;
2. Contributing to virtual reality and the emerging metaverse development by making VR authoring tools more accessible, as well as to the broader use and development of VR experiences, which can support various Sustainable Development Goals (SDGs);
3. The practical application and guidelines evaluation, not limited to proposing theoretical guidelines but also testing their practical application, providing a real-world evaluation of their effectiveness and demonstrating their operational reliability;
4. Highlighting areas for improvement in existing VR tools in terms of intuitiveness and user-friendliness.

This study evaluates the validity of fourteen design guidelines for the development of intuitive virtual reality authoring tools [11] by putting them to the test on an example tool: the NVIDIA Omniverse Enterprise, verifying qualitatively the use of this artifact in the stage depicted in green in Figure 1. NVIDIA developed the Omniverse, which seeks to shape the 3D internet and open metaverse by providing a platform for the development of industrial metaverse applications in the fields of robotics, scientific computing, architecture, engineering, manufacturing, and industrial digital twins [20].

The paper is structured into three sections: Section 2 outlines the materials and methods adopted, Section 3 examines the findings, and Section 4 offers our conclusions and recommendations for future studies.

2. Materials and Methods

The nature of our project is exploratory, because intuitive virtual reality authoring tools are a relatively under-researched area. This necessitates an initial deep dive to grasp and articulate the concept thoroughly, making qualitative research an ideal approach [21].

This study applied Design Science Research to contribute to scientific knowledge and practical application by addressing a research problem or opportunity [19].

Similar to the method used in prior DSR investigations [19,22], we followed the following six steps: (1) identify the problem; (2) define the solution objectives; (3) design and development; (4) demonstration; (5) evaluation; and (6) communication. Figure 2 outlines the key activities of each step.
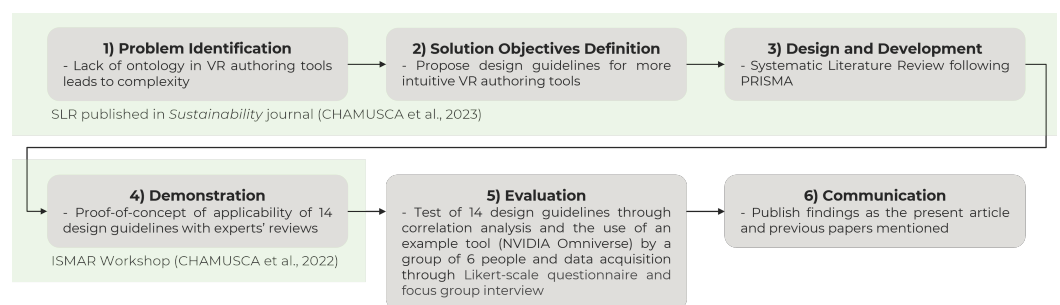

Figure 2. Design science research flow, adapted from Peffers et al. [22]. Steps 1, 2 and 3 described in [11], step 4 accomplished in [25].

The first three steps were completed by Chamusca et al. [11]. In Steps 1 and 2, the problem "lack of ontology in VR authoring tools leads to complexity" was identified and the solution objectives were set, which were to propose design guidelines for the development of intuitive virtual reality authoring tools. Step 3 involved doing a review of the literature in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) principles [23]. This was accomplished by employing a technique that comprises planning, scoping, searching, assessing, and synthesizing [24].

The outcomes of the literature review were synthesized, and the authors developed an artifact: the fourteen design guidelines described in Table 1.

Table 1. Design guidelines (DG) list, abbreviation code (AC), and frequent terms [11].

DG	AC	Frequent Terms
Adaptation and commonality	DG1	interoperability, exchange, data type, patterns, multiple, modular, export/import process, hardware compatibility
Automation	DG2	inputs, artificial intelligence, algorithms, translation, reconstruction, active learning, human-in-the-loop, neural systems
Customization	DG3	control, flexibility, interactions, manipulate, change, transformation, adapt, modify, programming, editing, modification
Democratization	DG4	web-based, popularization, open-source, free assets, A-FRAME, WebGL, deployment
Metaphors	DG5	natural, organic, real life, real-world, physicality, abstraction; embodied cognition
Movement freedom	DG6	manipulation, gestures, position, unrestricted, selection, interaction, flexible, free-form
Optimization and diversity balance	DG7	trade-off, less steps, fast, complete, limitation, effective, efficient, simplify, focus, priorities
Documentation and tutorials	DG8	help, support, fix, step-by-step, learning, practice, knowledge, instructions
Immersive authoring	DG9	what-you-see-is-what-you-get (WYSIWYG), engagement, 3D modeling, programming, 3D interaction, paradigm, creation, HMD
Immersive feedback	DG10	visual, haptic, hardware, multi-sensory, physical stimuli, senses
Real-time feedback	DG11	simultaneous, latency, WYSIWYG, synchronization, preview, immediate, run-mode, liveness, compilation, direct
Reutilization	DG12	retrieve, assets, objects, behaviors, reusable, patterns, store, library, collection, search
Sharing and collaboration	DG13	multi-user, multi-player, remote interaction, community, simultaneous, communication, network, workspace
Visual programming	DG14	primitives, logic, data-flow, nodes, blocks, modular, prototype, graphic

In Step 4, we demonstrated a proof-of-concept of the applicability of the proposed design guidelines by examining and amending them through skilled feedback, with early versions presented to researchers in conference sessions among them the Metaverse and Applications Workshop, held in the IEEE International Symposium on Mixed and Augmented Reality (ISMAR) [25]. The procedures used to complete the remaining 5 and 6 steps in this investigation are given below.

In Step 5, we evaluated the validity criteria for utilizing the fourteen derived design guidelines to determine the intuitiveness of current VR authoring tools on an example tool. We started this evaluation by applying the Pearson Correlation Coefficient (PCC) to the Chamusca et al. [11] results to find out how often two guidelines were found together in the studies that were looked at during the SLR (Section 3.1), since the developed guidelines complement one another and were not given separately. In a later step of this study, this analysis was used along with the Likert-scale questionnaire scores as another indicator to evaluate the validity of the design guidelines (Section 3.2). It was expected the questionnaire scores given to the design guidelines would match the correlation results.

Still in Step 5, we conducted an experiment with six engineering students from our Extended Reality for Industrial Innovation and Sustainable Production Lab (referred to as participants P1–P6). Following Creswell's [21] guidelines for qualitative research, we purposefully selected a small and focused group of participants since qualitative research aims to choose participants that offer the most insightful understanding of the research question, rather than aiming for a broad, random sample typical of quantitative studies. Our choice of six participants aligns with these guidelines, aiming to provide depth and clarity to our exploratory study without the need for a large sample size.

They had no background in programming, no prior experience using the exemplary authoring tool or creating virtual worlds in any form of game engine, and no prior awareness of the fourteen design guidelines. However, some of them mentioned a basic understanding of 3D modeling and/or navigation.

Participants in the study assessed design guidelines using the NVIDIA Omniverse Enterprise package as an example authoring tool (https://www.nvidia.com/en-us/omniverse/, accessed on 10 December 2022). Although NVIDIA has not specifically indicated so, for the purposes of this study, the Omniverse components are regarded as an authoring tool, as just a subset of its available tools were utilized in our experiment, including only the most relevant features for content creation. The evaluation of a tool is part of its life cycle and, consequently, enters the process of product design and may generate improvements to be implemented. Therefore, the design guidelines must work as a reference for the whole software product design process, including their evaluation (Figure 1).

We chose NVIDIA Omniverse as a use case because it helps create virtual worlds and the metaverse through virtual collaboration, 3D simulation, modeling, and architectural design [16,26]. The Omniverse's main features include virtual reality, artificial intelligence to analyze audio samples and match them with meta-humans' facial animation, 3D marketplaces and digital asset libraries, connectors to outside applications like Autodesk Maya (https://www.autodesk.com/products/maya, accessed on 10 December 2022) and Unreal Engine, and the integration of 3D workflows like digital twins [27]. The platform was used, for example, to build a digital twin for BMW that improved the precision of its industrial work by combining real-world auto factories with VR, AI, and robotics experiences [28].

Industrial concerns are gaining a lot from the engineering simulation available on this tool, even though it was the creative sector that gave virtual worlds their initial impetus through game development and entertainment studios [29]. For professional teams, NVIDIA Omniverse Enterprise can develop comprehensive and photo-realistic design platforms that enable better designs with fewer expensive mistakes in less time. Teams of designers, engineers, marketers, and manufacturers can work together through the Omniverse Nucleus Cloud. This lets creators in different places share and collaborate in real time on designing 3D scenes for industrial applications, like car design [16,27,29]. However, even similar to VR authoring tools in the professional context, Omniverse can be seen as complex for having many sub-products, requiring some time to learn the user interface, presenting a challenge to find the most efficient way to use it, and requiring research in its documentation [30].

The six participants experimented the hands-on lab *Build a 3D Scene and Collaborate in Full Fidelity* (Figure 3a) taking turns with three NVIDIA LaunchPad free tryout accounts for Omniverse Enterprise (Figure 3b). LaunchPad gave users access to NVIDIA virtual machines with graphics capabilities that they could use to run Omniverse apps like Create and View.

Figure 4a illustrates the experiment scope, limited to the activities described on the topics *Overview*, *Step #1: Setting Up Your Environment*, *Step #2: Start Creating*, *Designer #1*, *Designer #2*, and *Designer #3*. Each one of these topics provided documentation and videos that could guide the participants through the activities, therefore no prior training was needed. An onboarding meeting was made share the licenses credentials with the participants, divide them into two groups and set the day and time for their sessions. Afterwards they used the LaunchPad documents to learn how to install the virtual machine

application, and run the Create application, and work together to build a 3D scene of a park, as shown in Figure 4b. The same scene could be seen by three people at the same time, each using one of the three accounts that had been requested before. Each participant should execute the activity described in the topics entitled *Designer #1*, *#2* or *#3*. The activities included adding an environment, adjusting lighting, adding 3D assets from a library, adding or changing textures from a library, and organizing the work layers to guarantee the organization of the space while also avoiding the conflict of more than one person editing the same object at once.
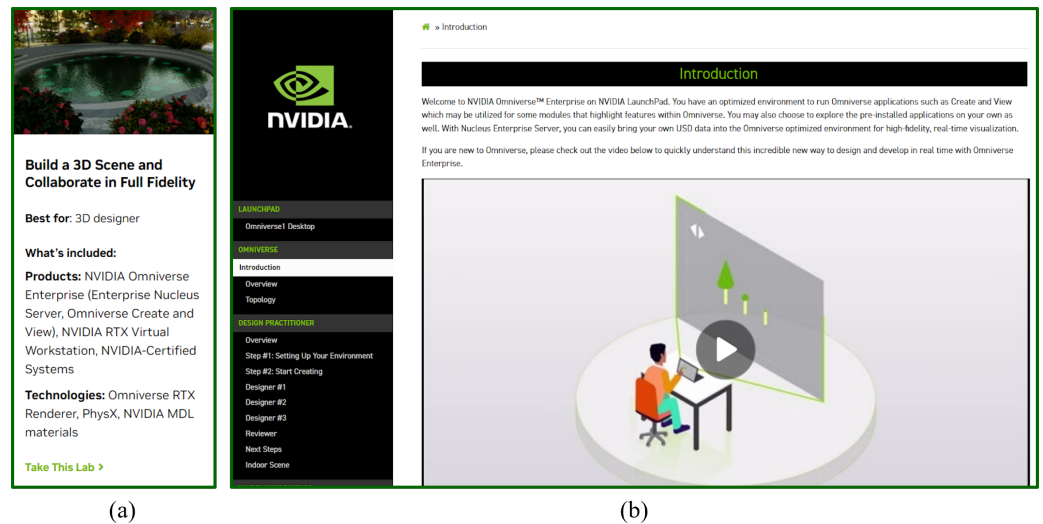

Figure 3. (**a**) Hands-on lab; (**b**) NVIDIA LaunchPad interface for Omniverse Enterprise.
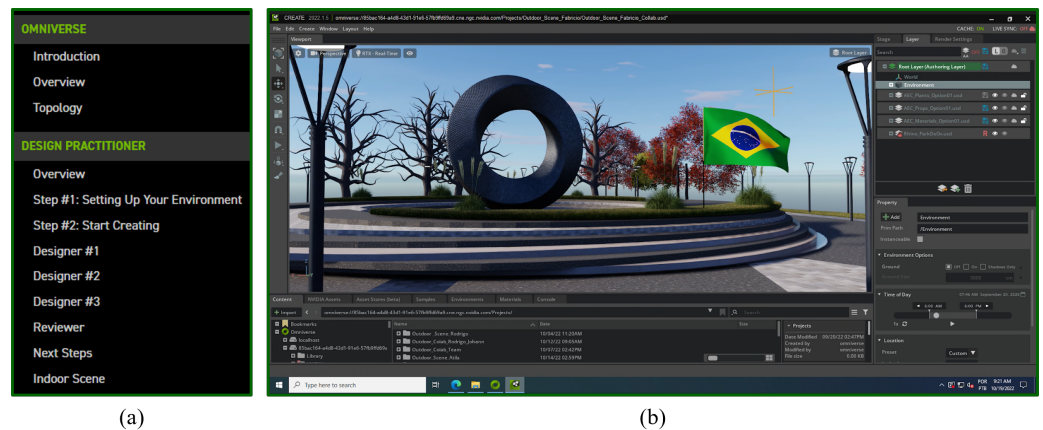

Figure 4. (**a**) Tutorial steps for the Create platform; (**b**) screenshot of the Create interface.

Before using LaunchPad to get into Omniverse, the participants read a document that explained each design guideline in detail [11]. Then, we answered questions for further clarification on the design guideline definitions. After that, the participants took turns using the accounts. All Omniverse LaunchPad sessions were done through online remote meetings.

Then, we captured the participants' insights about the design guidelines using two data acquisition methods. The first method was a Likert-scale questionnaire comprising fifteen questions (Supplementary Materials), which they answered individually in the format of an online forms. The scale had a numeric scale that ranged from *totally disagree* (1 point) to *totally agree* (5 points), which should be marked according to their agreement about the existence of a design guideline in Omniverse. Additionally, the participants were questioned about significant observations made throughout the execution of the

tutorials, which could include system errors, challenges, and interesting functionalities. The equations shown in Figure 5 are a first proposal of how to estimate a punctuation to an authoring tool's intuitiveness using the guidelines, where the *guideline score* corresponds to the average of participants' answers on the Likert-scale questionnaire (1–5) and the *final score of the tool* evaluated stands for the sum of all guideline scores. These equations were applied to the experiment realized in this study so the answers obtained with the second method could be compared to other indicators.

$$\overline{G} = \frac{\sum_{i=1}^{p} G_i}{p} \qquad\qquad F = \sum \overline{G}$$

(a) $\qquad\qquad\qquad (b)$

$\overline{G}$ = *guideline score*
G = *participant answer in Likert−scale*
p = *number of participants*
F = *final score of the tool*

Figure 5. (**a**) Average of participants' answers on the Likert-scale questionnaire (1–5); (**b**) sum of all guideline scores.

The number obtained as the final score was compared to the maximum score value in the questionnaire, which is equal to 70, considering the product between fourteen guidelines and five points for *totally agree*. It was assumed that a percentage lower than 50% of this total value would characterize authoring tools that are not very intuitive, while a higher percentage would indicate greater intuitiveness. The questionnaire results (Section 3.2) were also matched to the correlation analysis results, using PCC, (Section 3.1) to confirm the similarities, which were determined by examining the score of the guidelines with strongest positive and negative correlation obtained on the questionnaire. It was expected that guidelines with strong positive correlation would receive similar scores and, therefore, lower difference values, while guidelines with strong negative correlation would receive very different scores and, therefore, higher difference values. However, these values were obtained to serve as a demonstration of how the guidelines could be used to evaluate a VR authoring tool and to be compared with the results obtained with the second method. It was assumed that, if the results obtained from testing the tool were compatible with expectations, this would be a good indication of the validity of the guidelines as a guide to develop and evaluate more intuitive VR authoring tools.

The second data acquisition method was a focus group interview (Section 3.3), in which participants answered eighteen questions on their understanding of the design guidelines and their experience using them to evaluate the exemplary use case (Supplementary Materials). The answers were collected in an in-person group meeting with a duration of two hours by recording audios using a smartphone and previously converting them to text using an online tool, which was then analyzed in the results session. Finally, we provide a pipeline including a compilation of all the steps carried out in this study, as a guide for anybody wishing to replicate the experiment using different VR authoring tools. Figure 6 contains a summary of the experiment conditions, with the start point steps, the experiment context, and the data acquisition methods.

Step 6 involves disseminating the study's results through this publication, in which we demonstrate how an evaluation experiment using a VR authoring tool may be undertaken from the perspective of the design guidelines and, therefore, assessing the validity of the guidelines as an artifact.
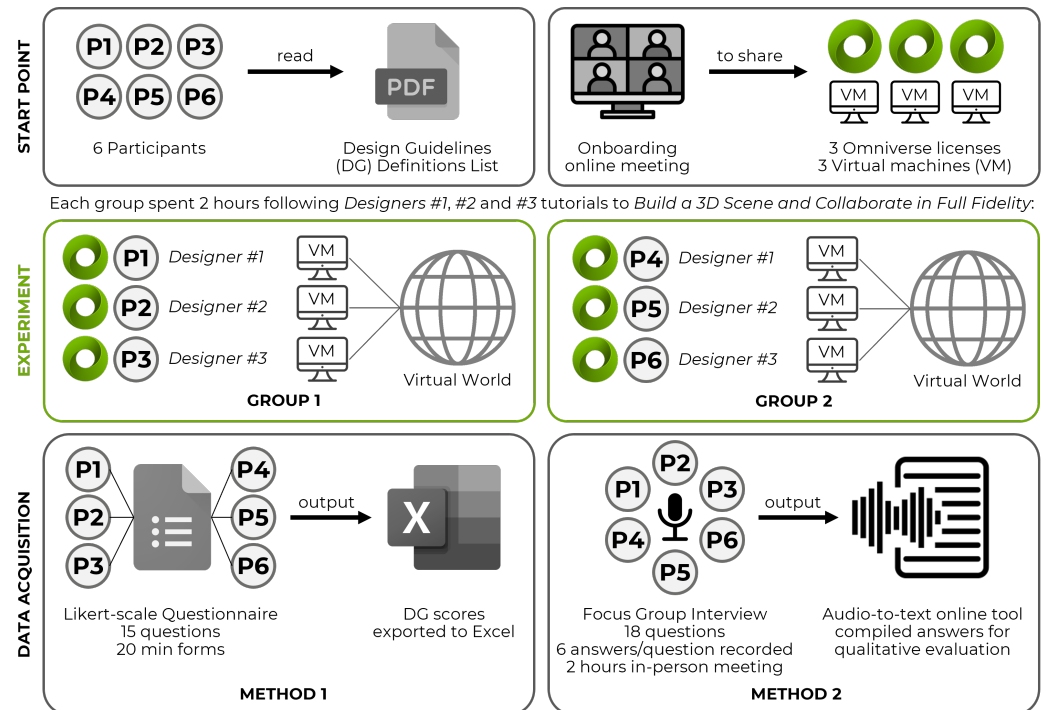
Figure 6. Summary of experiment conditions.

3. Results

In the following sessions, we describe our findings.

3.1. Reviewing the Design Guidelines

The majority of the authoring tools determined by the review are proof-of-concept, but the design guidelines may promote the emergence of commercial platforms with fewer constraints, making accessible technology and enhancing its stage of development. Furthermore, the results reported by Chamusca et al. [11] help with starting or expanding the establishment of ontologies for facilitating the creation of virtual reality authoring tools related to the identified gap [12]. The shortage of ontologies associated with virtual reality authoring tools suggest that there are few associated conventions governing the creation of these platforms [12].

Furthermore, guidelines can positively assist the establishment of the metaverse by simplifying the application of its essential components. The large scale of this notion leads to scarce knowledge of how it operates, demanding a taxonomy for the metaverse [31]. The advised taxonomy acknowledged the following components as critical to the metaverse's emergence: hardware, software, and content. Chamusca et al.'s design guidelines [11] have several parallels with technologies that have lately become issues in the metaverse and were mapped as hardware, software, and content [31].

Chamusca et al. [11] define *intuitiveness* as being able to accomplish tasks effortlessly with little direction; lowering the barrier to entry; requiring less information, time, and fewer steps; being suitable for both expert and non-expert users; being aware of and feeling present in virtual reality; feeling comfortable with the tool; making few mistakes; and using natural movements. While there is no specific way for measuring intuitiveness, aspects like usability, effectiveness, efficiency, and satisfaction may be examined with regularly used questionnaires like the After-Scenario Questionnaire (ASQ) and the System Usability Scale (SUS) [12,32].

The utilization of questionnaires as a well-established method to evaluate software tools was the source of the idea of using the guideline artifact in association with a questionnaire to help the process of evaluating virtual reality authoring tools. This is also supported by the contribution of the guidelines to the creation of ontologies and taxonomies in the

field. Standard concepts, methods, and terminology are absent in the creation of VR authoring tools with vastly diverse types, as is the usage of multiple assessment approaches to determine their usability [12].

The developed guidelines complement one another and were not given separately [11]. Figure 7 shows the correlation analysis that was performed with the fourteen design guidelines. It shows which pairs of guidelines show up together more or less often in the works that were reviewed. The three strongest negative and positive correlation values (CV) in Figure 7 were captured, and from that, the pairs of design guidelines that presented these values were highlighted in Tables 2 and 3. The columns related to questionnaire scores (QS) link the scores for each guideline presented in Figure 8, which will be explained in more detail in Section 3.2, to the correlation analysis.

Examining the cases of Democratization (DG4) and Adaptation and commonality (DG1), the strong positive correlation can be associated with the fact that multiple elements related to DG1 can, consequently, lead to DG4. For instance, utilizing the same authoring tool on multiple systems and allowing different file types for the same kind of data might assist optimize and expand a tool's availability. Movement freedom (DG6) and immersive authoring (DG9) are interdependent, since DG6 cannot occur without DG9, while the contrary is possible. In fact, DG6 complements DG9 by underlining the necessity of movement freedom throughout an immersive authoring process. Metaphors (DG5) can be used to convey movement freedom (DG6), such as sliding and arranging items as if they were in the actual world and linking objects that are physically separated by drawing apparent paths between them.
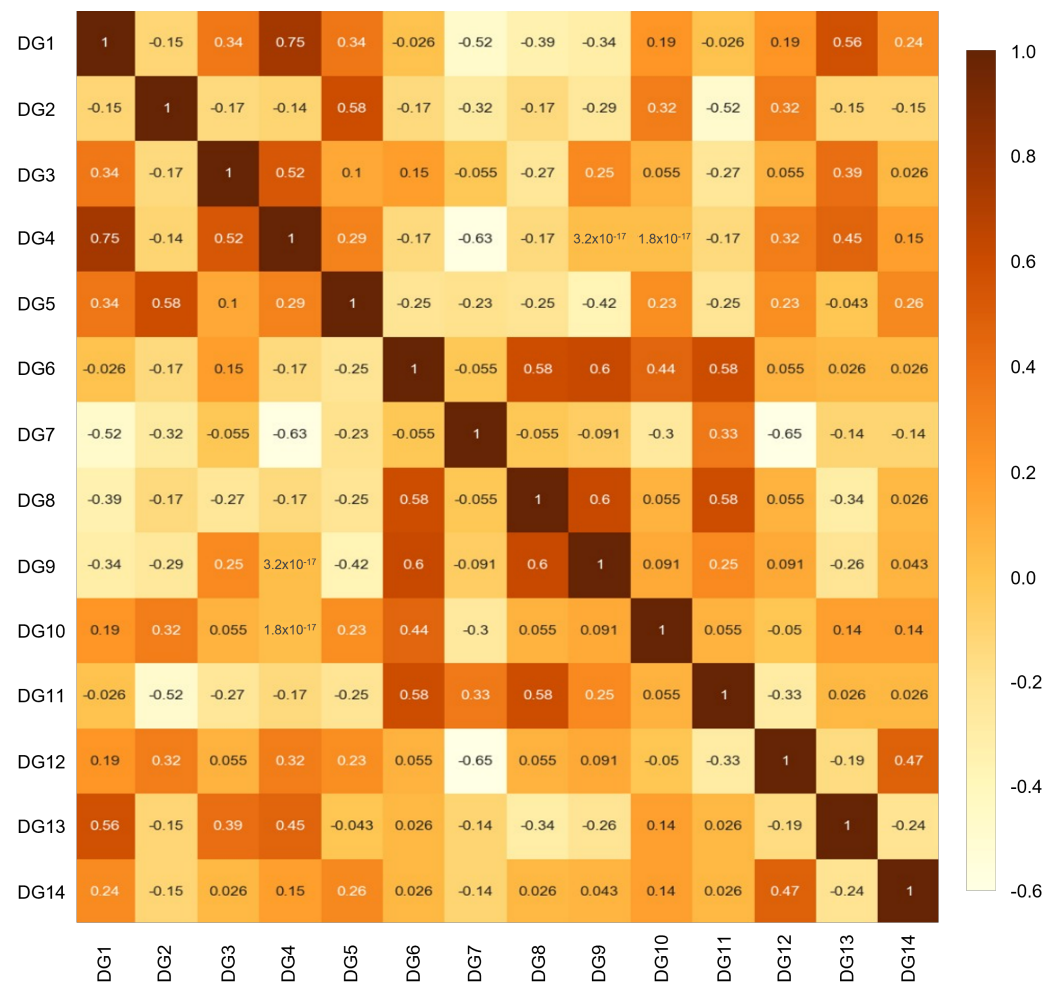

Figure 7. Applying the Pearson Correlation Coefficient (PCC) to the fourteen design guidelines.
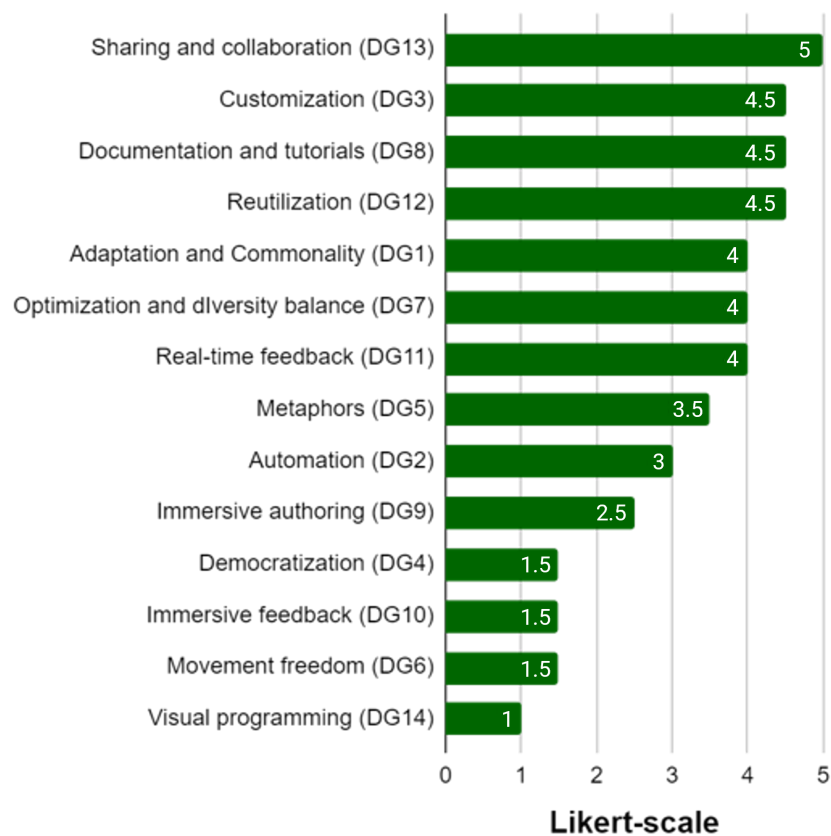
Figure 8. Average value of each guideline's determined score for the exemplary use case.

Table 2. Design guideline pairs with the strongest positive correlation.

CV	Design Guidelines Pairs	QS	QS Dif.
0.75	Democratization (DG4) and Adaptation and commonality (DG1)	1.5 (DG4) and 4 (DG1)	2.5
0.60	Movement freedom (DG6) and Immersive authoring (DG9)	1.5 (DG6) and 2.5 (DG9)	1
	Movement freedom (DG6) and Metaphors (DG5)	1.5 (DG6) and 3.5 (DG5)	2
0.58	Documentation and tutorials (DG8) and Automation (DG2)	4.5 (DG8) and 3 (DG2)	1.5
	Metaphors (DG5) and Immersive authoring (DG9)	3.5 (DG5) and 2.5 (DG9)	1
	Real-time feedback (DG11) and Immersive authoring (DG9)	4 (DG11) and 2.5 (DG9)	1.5
	Real-time feedback (DG11) and Metaphors (DG5)	4 (DG11) and 3.5 (DG5)	0.5

Table 3. Design guideline pairs with the strongest negative correlation.

CV	Design Guidelines Pairs	QS	QS Dif.
−0.65	Immersive feedback (DG10) and Reutilization (DG12)	1.5 (DG10) and 4.5 (DG12)	3
−0.63	Immersive feedback (DG10) and Democratization (DG4)	1.5 (DG10) and 1.5 (DG4)	0
−0.52	Immersive feedback (DG10) and Adaptation and commonality (DG1)	1.5 (DG10) and 4 (DG1)	2.5
	Real-time feedback (DG11) and Automation (DG2)	4 (DG11) and 3 (DG2)	1

Documentation and tutorials (DG8) are frequently generated utilizing Automation (DG2), such as AI assistants that identify when a user faces difficulties with a task and make clever advice to solve it. Metaphors (DG5) can make Immersive authoring (DG9) simpler by transforming conceptual ideas into concrete implements, such as using controller buttons

to replicate steps like the ones we would perform in actual life, such as pulling the trigger button to grab an object and letting go it for dropping it. Ultimately, Immersive authoring (DG9) and Metaphors (DG5) require Real-time feedback (DG11) to perform effectively. This enables authors of content to have a *what you see is what you get experience*, which means the user truly views the virtual scene while composing it [11].

Regarding the design guidelines with strong negative correlation, it is remarkable that Immersive feedback (DG10) appears on three of the four correlations. This makes sense, because the definitions brought by DG10 are really unique for the immersive context, making the use of some kind of virtual reality device mandatory. Reutilization (DG12), Democratization (DG4), and Adaptation and commonality (DG1) are not guidelines limited by the use of devices, being more generalist to the virtual world creation. Safe conduct, Adaptation and commonality (DG1) could indirectly contribute to Immersive feedback (DG10), considering that allowing communication with different types of VR hardware is one of its definitions. Real-time feedback (DG11) and Automation (DG2) are two guidelines connected to a good system infrastructure, and automated functions should have real-time feedback but nothing more than that.

The correlation between guidelines are a comparison indicator that serves as a starting point for judging the intuitiveness of an authoring tool. It is expected that the guidelines receive a score consistent with the way they correlate and when this does not happen, interpretations can be made to lead to a future improvement in the tool's intuitiveness. These results illustrate that it is possible to assess the existence of guidelines on a tool by understanding how they relate to one another, resulting in an indicator to evaluate the design guidelines artifact, which were carried out in Section 3.2.

3.2. Likert-Scale Questionnaire

After executing the tutorial described in the NVIDIA LaunchPad, the participants answered the Likert-scale questionnaire, followed by the detailed document about the design guidelines. Figure 8 presents these answers, with the design guidelines ranked by the average value of their scores, as determined by the equation provided in Figure 5a.

The five guidelines with higher scores are shown in the following topics with examples of where the guidelines were seen by the participants, according to their comments:

1. Sharing and collaboration (DG13): the participants could see in real time the updates made by the others, and they finished the activities quicker by splitting the job between more people;
2. Customization (DG3): the participants could easily change the color and texture of the assets imported from the libraries;
3. Documentation and tutorials (DG8): the LaunchPad itself promotes a good step-by-step for a first try of the tool, giving an enough number of activities so the person can get to know the tool without being lost in numerous tutorials;
4. Reutilization (DG12): Omniverse Create has libraries of assets with many 3D models and textures available, so the participants did not need to look for them outside the software;
5. Adaptation and commonality (DG1): the participants could see the same file being updated in real time on the Omniverse View, while the scene was being created on Omniverse Create; also, the asset libraries were integrated with the software interface, so they did not need to worry about file extension compatibility or do an extra process to import them.

The five guidelines with lower scores were: Immersive authoring (DG9), Democratization (DG4), Immersive feedback (DG10), Movement freedom (DG6), and Visual programming (DG14). We could not run a test using virtual reality during the experiment with the exemplary use case because the NVIDIA LaunchPad did not provide the tool Omniverse XR, which certainly caused the decrease in the score given to the guidelines related to immersiveness, which are Immersive authoring, Immersive feedback, Movement freedom, and Visual programming. This demonstrates that the participants understood

the design guidelines' definitions since, even though they are not specialists, they were able to understand that the experience did not fit their descriptions and disagreed with the presence of these guidelines.

We also observed that, similar to the complex game engines frequently used for VR development today, such as Unreal and Unity, in the version of Omniverse Enterprise experimented as the exemplary use case, virtual worlds for VR experiences are still developed primarily using 2D screens, not HMDs and other wearables. This is different from what Chamusca et al. [11] saw during the development of the guidelines, since the reviewed works showed that adding virtual reality equipment to the process of creating an VR experience can make it easier to understand and do it correctly. This indicates that the guidelines were comprehensible and the participants did not perceive intuitiveness in creating an immersive experience without being allowed to test it along the way.

Democratization (DG4) was at the bottom of the list, probably because Omniverse Enterprise is not free and can only be used with paid NVIDIA accounts or limited free tryout accounts, which were the case in this study. Also, technical problems related to the high latency of the virtual machines faced by some participants probably affected the results, which will be discussed in the next Section 3.3. On the other hand, all the participants could complete the activities proposed in the exemplary use case, even though they had never used similar software before.

Using the equation shown in Figure 5b to calculate the sum of all guidelines scores and comparing them to the maximum score value in the questionnaire, we obtained a total score of 45 out of a maximum of 70, or 64%. This percentage represents the global level of intuitiveness of a VR authoring tool from the guidelines' perspective, as experienced by the participants while executing the experiment. This average score is aligned with the declaration that the Omniverse tool can be seen as complex, requiring time to understand the user interface, presenting a challenge to find the most efficient way to use it, and requiring research in its documentation [30]. This contributes to the validity of the design guidelines since the medium score of 64% obtained from their perspective, matching past feedback about the software.

Regarding the correlation between the guidelines, most of them were in line with the results shown in Section 3.1 when the difference between their scores was checked. It was assumed that guidelines with strong positive correlation values would have lower difference values, while those with strong negative correlation values should have high difference values. Tables 2 and 3 show that the design guideline pairs with strong positive correlation values had a score difference of around 0.5 and 1.5, while the pairs with strong negative correlation values had a score difference of around 2.5 and 3, which matched the expectation. However, an unexpected score difference of 2.5 in Table 2 and 0 in Table 3 draws attention, having the guideline Democratization (DG4) as a common factor.

This indicates that some unexpected occurrence connected to the Omniverse experience produced a mismatch between this guideline and the others, most likely the same incident that led to this guideline's low score on the Likert-scale questionnaire. During the focus group interview, which will be discussed in Section 3.3, participants talked about problems like *the program taking too long to respond to commands* and *difficulty installing the virtual machine*. Such problems are not directly related to the usability of the tool, but rather to the specific circumstances of each participant, such as an incompatible internet connection. This may have caused a decrease in the Democratization (DG4) score to 1.5, not following the expectation of having a higher score such as Adaptation and Commonality (DG1) with 4 points, with which has a strong positive correlation of 0.75, leading to the high score difference of 2.5. Technical issues in conjunction with the absence of Omniverse XR approximated the Democratization (DG4) score with the low results of the immersiveness-related guidelines, Immersive feedback (DG10) being one of them with 1.5 points, with which DG4 has a low correlation level of -0.63, but a low difference score of 0 in this experiment.

3.3. Focus Group Interview

The participants' responses obtained with the focus group interview are examined in the following section.

3.3.1. The Exemplary Use Case Omniverse Tool

During the execution of the experiment, the participants encountered both obstacles and opportunities associated with the activities proposed in Omniverse LaunchPad. Four of the participants said that applying textures to small areas was the hardest part. This includes actions applying grass on a small piece of the 3D ground mesh. Three participants said that the software took too long to respond to commands, which could be caused by technical problems like incompatible internet connection.

Only one participant mentioned difficulty starting the program and following the LaunchPad step-by-step instructions for installing the virtual machine. Two participants had difficulties understanding how to navigate inside the 3D environment, which includes rotating the camera and zooming in and out on objects, while two other participants considered this an easy and intuitive task.

> *"There was a step where it was asked to apply grass to a specific area, and I was not able to select it. That step really stuck with me. I did not know if it was because I was not using the right tool, if I had skipped a step, or if the tutorial was not able to instruct me to reach my goal."*—P2

Omniverse LaunchPad provided links to external videos along the tutorial with more details on some features, such as applying textures to meshes. Possibly, participants who had difficulty with this function did not notice these links in the explanation or limited themselves to only follow the instructions on the main page with the activities. Four participants said that importing 3D assets from the Sketchfab (https://sketchfab.com/, accessed on 10 December 2022) library and placing them in the scene was one of the easiest things to do. Another participant highlighted how easy was to set the environment's illumination for the skybox using a slide button that changed the position of the sun in real time.

> *"For me, the most intuitive part was adding and removing objects such as trees, vases, benches, and tables. It was very intuitive because it has an illustrative icon of what you were going to place, and when you select and drag it into the scene, the software tells you where that object will be, and you can even see it from different perspectives."*—P4

When asked if they had already used a tool similar to the exemplary use case Omniverse, three participants mentioned they had already had contact with parametric 3D modeling software (Solidworks https://www.solidworks.com/, accessed on 10 December 2022), three cited games like The Sims and Minecraft as facilitators, and only one had already had a brief contact with a game engine (Unity) but with the intention of create a 2D mobile application. We found that participants who had previous experiences with software or games that required interaction and movement in a 3D environment found Omniverse easier to use because the controls are usually very similar.

> *"I had not used a tool like Omniverse before, but something that might have made it easier was my experience with games like The Sims, as you created an environment and inserted objects."*—P3

3.3.2. Guidelines Identification

Along with the activity to be carried out for the exemplary use case Omniverse, participants were provided with a detailed document describing the fourteen design guidelines' definitions and a Likert-scale questionnaire that asked if they agreed, or disagreed, with the presence of the guidelines in association with the software functions used in the activity. To efficiently answer the questionnaire, most of the participants (four) chose to take notes as they followed LaunchPad tutorials, using the guidelines' document as a support during this process. Only two participants did not take notes, although they did consult the guidelines' definitions in order to be able to answer to the questionnaire coherently.

Despite being instructed to identify the presence or absence of the design guidelines in the tool under test, the participants were not told how to do it. When asked about their method for associating the guidelines with Omniverse, the participants answered that they focused on identifying the steps they found complex or easy to accomplish and connecting them with the definitions of the guidelines. Most did it in a segmented way, i.e., after completing each step instructed by LaunchPad, so that all the details were clear in their memories. Another way of highlighting the presence or absence of a guideline in the experimental tool was the association with the examples given in the definitions of the guidelines; if an example was directly found, positive points were given to the guideline.

Participants also mentioned that some guidelines were obvious while others required more reflection, particularly on whether their presence or absence would be limited to a specific stage of the activity or was truly part of the Omniverse's characterization as a tool. Among the guidelines that were easier to identify were Automation, Customization (cited three times), Democratization, Movement freedom, Documentation and tutorials (cited twice), Real-time feedback (cited twice), Reutilization, Sharing and collaboration, and Visual programming (cited twice). Listed below are statements from participants that demonstrate their reasons for identifying these guidelines as easy to identify:

- *"In group dynamics and collaboration, I could see the almost instantaneous change of material, color, or movement made by other people"*—(P1, about Real-time feedback);
- *"This guideline did not exist, and because of that, I had a lot of difficulty with the slowness to perform some actions"*—(P2, about Real-time feedback);
- *"I pointed out this guideline because I could not find it during the experiment, so it was very easy to identify"*—(P3 and P6, about Visual programming);
- *"I was impressed with what a person is able to do using Omniverse through a virtual machine accessed by a mere notebook, since even using a computer with a good GPU, the graphics processing of programs like this takes a long time"*—(P6, about Democratization);
- *"The tool has a library with assets you can place and reuse in the environment"*—(P3, about Reutilization).

Among the guidelines considered more difficult to identify, the following were mentioned: Metaphors, Movement freedom (cited twice), Optimization and diversity balance, Immersive authoring (cited twice), Immersive feedback, Sharing and collaboration, and Visual programming. Below are some of the participants' statements that show their motivations for pointing out these guidelines as difficult to identify:

- *"I had a lot of difficulty answering the question about this guideline. I had to read its description several times to find out if the LaunchPad would apply with the definition"*—(P1, about Immersive authoring);
- *"Even interacting with an open environment, I felt a little limited, so I kept questioning whether I really had this movement freedom or if it was a freedom within the limitation of using the software through a 2D screen"*—(P1, about Movement freedom);
- *"I found it a little subjective; I could not say to what extent we can consider that the process was optimized or not, and whether it was complex or not"*—(P2, about Optimization and diversity balance);
- *"The most difficult for me were the two that involved immersion, because I believe it is subjective to identify if I am immersed in that environment; what may be immersive for me*

> *may not be immersive for someone else, and vice versa"*—(P3, about Immersive authoring and Immersive feedback);

- *"I had to read the guideline a few times to have a better understanding when answering, due to my lack of knowledge in the area"*—(P4, about Visual programming);
- *"I could not say if that was easy or not, because I did not have much experience with collaboration in other similar applications and software, so Omniverse collaboration might not be efficient in front of the guideline"*—(P5, about Sharing and collaboration).

Guidelines classified as features or requirements were equally mentioned as easy or difficult to identify, so no discussion can be given on that. However, the Movement freedom, Sharing and collaboration, and Visual programming guidelines were mentioned both as easy and difficult to identify by different participants, which may represent ambiguity in the definitions given to them and, consequently, a lack of standards to determine situations in which these guidelines apply or not. This was clear from what the participants said, since they were not sure about the meaning of some of the terms used in the guidelines' definition. Immersiveness, for example, was not directly linked to virtual reality experiences by the participants, but all of the examples in the definition of the guidelines are linked to this aspect. This can also be attributed to the participants' lack of experience with the area and its technical terms.

The lack of experience may also be the reason why the guidelines with highest scores on the Likert-scale questionnaire (Sharing and collaboration, Customization, Documentation and tutorials and Reutilization) were presented as easy to identify, while four of the guidelines with the lowest scores (Immersive authoring, Immersive feedback, Movement freedom and Visual programming) were presented as difficult to identify. This suggests that even though the participants were able to discern that the low score guidelines were not featured in the tool, they still had doubts when responding to the questionnaire, indicating that they were challenging to recognize. The inverse is true of the guidelines with the highest scores, which were easily observable throughout the execution of the experiment and could, thus, be better evaluated. This raises the question of whether the difficult-to-identify guidelines had subjective descriptions, as many of the participants claimed, or whether the fact that the tool did not present the examples stated by its definitions led to a lack of clarity for the interpretation of the participants, who were unable to implement the concepts illustrated in the examples.

In this perspective, the Democratization guideline stands out because, unlike the others with low scores on the questionnaire, it was presented as easy to identify, preserving the history of inconsistencies revealed throughout the experiment. Comparing Democratization's score to the correlation analysis revealed unexpected findings, which could be attributed to the fact that the tool is not free and that technical issues occurred throughout the test. Given that not all participants experienced technical difficulties during the experiment, P6's generally positive comment in this Section may add to the prior claims. In addition, the fact that a guideline is considered easy to identify should not be correlated with its presence, as participants P3 and P6 made evident in their comments regarding the absence of Visual Programming.

3.3.3. Guidelines Strengths and Weaknesses

Then, the participants were asked about the strengths and weaknesses related to the use of guidelines for evaluating the intuitiveness of existing authoring tools for experiences in virtual reality (Figure 1). Three participants said that the inclusion of practical examples to the description of the guidelines was the greatest strength. This was due to the fact that the examples made it feasible to compare the assessed tool functionalities to those of other software or apps throughout the experiment, despite the fact that part of the general description was not very clear. Moreover, titles were cited as strengths, since they allowed for rapid reference to what the guideline defines.

> *"I think the titles were very striking and helped us understand what that guideline meant. See Real-time feedback, for example. Just reading the title, I can easily associate it with the definition without necessarily having to read it."*—P2

Participants pointed out that one of the weaknesses was the use of unusual words like *haptic*, that were derived from the field's technical terminology. Other examples were the subjectivity of some of the definitions, the lack of visual references, such as pictures, to compose the definitions of the guidelines, and the lack of delimitation to make more clear the difference between guidelines with similar names.

> *"Some guidelines, such as Metaphors, are very subjective, which could be solved using images, for example."*—P6

Concerning the presented set of guidelines, all participants agreed that it was appropriate and complete. They did not suggest any additional guidelines to be added to the list, although some believe that as technology evolves, new guidelines may be necessary.

According to all participants, the guidelines have different weights in terms of intuitiveness. This indicates that the presence of guidelines with a higher weight makes a tool more intuitive, whereas those with a lower weight have less of an effect. However, there was no consensus among the participants about which guideline would have higher or lower weights, so this topic should be treated as a future research. Three of the participants stated that the relevance of the guidelines varies based on the context in which a tool is being assessed. For instance, if the experience is collaborative or individual, or if the technology includes head-mounted displays and other VR peripherals, the relevance of certain guidelines changes.

> *"I believe the guidelines have different weights. For example, I consider Democratization to have a high weight in terms of intuitiveness, whereas Visual programming I consider to have a lower weight when analyzing a tool."*—P6

All participants believed that most of the guidelines were self-explanatory. However, some of them are subjective, making it difficult to use them to evaluate VR authoring tools, as their existence or absence can be understood differently by each individual. Even so, all participants stated that they would use the guidelines to assess the intuitiveness of alternative VR authoring tools. This is due to the fact that the guidelines helped them comprehend the potential of using Omniverse, and how it could be implemented. One of the participants believes that using the guidelines to evaluate other authoring tools will also contribute to the improvement of their definitions. Two others said that the guidelines can assist them in finding a tool that satisfies the requirements for the development of a particular project.

> *"The idea is to use the guidelines to find the tool that best meets the requirements of your project."*—P5 and P6

3.3.4. Changing Suggestions for the Guidelines' Future

In an effort to improve the concept of the guidelines, participants were requested to suggest changes and future applications. The majority of proposed modifications involved rearranging and categorizing the guidelines, including, for instance, a reduction in their number and convergence of those with comparable concepts. In order to guide the evaluators to assess an authoring tool through a certain sequence of the guidelines list, it was suggested that the guidelines be reorganized into those to be judged before testing with a tool and those to be judged during the experiment. Moreover, the parameters might

be categorized as applicable to the evaluation of 2D experiences, virtual reality immersion, or both. In the end, one participant disagreed with the suggestions to make modifications because he believed it was essential to analyze each guideline as it is now written.

> *"I believe you can reduce the number of guidelines by grouping concepts that are similar together; for example, Metaphor and Movement freedom are very similar in my interpretation."*—P1
>
> *"I believe Democratization should be one of the first guidelines to look for before experimenting with a tool because, if the tool does not have it, a much smaller audience will be able to access it and really evaluate the following guidelines."*—P2

For future implementations of the guidelines, the participants proposed replicating this experience, primarily by altering the composition of the evaluation group and the software tools evaluated. For instance, the application might be conducted with a group of industry specialists, such as programmers and VR experience designers, in order to obtain more technical input, since they are also the target audience for the guidelines application as a development guide for new VR authoring tools. The present investigation selected a group of participants with different degrees of experience, which may have led to variations in scores and interpretations of the guidelines' principles. The same test can be administered to individuals of different generations, such as children, teenagers, and the elderly, in order to compare their findings based on their technological experiences.

Participants also suggested conducting more extensive testing with each of the guidelines individually, examining specific experiences to identify them in tools, and then returning to the test collectively. About altering the software tools evaluated, identifying those that are recognized as intuitive on the market can help to confirm whether or not the guidelines are effective, since high scores would be expected. Reproducing the experiment using a tool that serves a different purpose or in a situation that enables the experience not only on 2D screens but also on head-mounted displays may illustrate that the guidelines are applicable to a wide range of authoring tools.

> *"I think the next step would be to test other authoring tools, especially those that allow authoring in virtual reality. We tested a tool mainly for 2D editing during this experiment, perhaps immersed in virtual reality we will have other insights that we have not noticed yet."*—P6

This leads to a discussion of the consequences of not being able to utilize head-mounted displays during the current experiment. Even though they knew what the immersiveness guidelines meant, all of the participants reported that it was difficult for them to evaluate the tool based on these guidelines. If everyone had tested the tool in virtual reality, their responses about Immersive authoring, Immersive feedback, Movement freedom, and Metaphors would be different. Nonetheless, the majority of them took this into account when answering the questions. Figure 8 demonstrates that these recommendations earned low scores.

All participants were aware that, in the context of the experiment, the example use case Omniverse lacked immersive elements, which resulted in a lower score. This demonstrates the effectiveness of the guidelines for the evaluation of existing VR authoring tools. In addition, even though the intuitive creation of virtual reality experiences is the final objective of the design guidelines, a significant portion of this creative process consists of developing the virtual world on 2D screens. Yet, the literature review indicates that the incorporation of virtual reality devices throughout the creation of the experience makes the process more intuitive and straightforward to implement, since the author will have the same experience as their audience along the way.

> *"Although not having had the experience of immersion in VR, it was not difficult to judge Omniverse in compliance with the guidelines. However, I disagreed with the existence of the guidelines related to immersion because I did not live the experience and hence did not recognize it in comparison to what I saw in Omniverse."*—P5

3.3.5. Further Considerations

Throughout the experiment, the Internet connection, the configuration of the virtual machine, and the execution of some software operations presented technical issues or took too long for certain participants. The participants were asked if these concerns affected their overall impressions of the experiment. Three participants claimed that they did not encounter any technical issues or that the issues were minor and had no effect on their performance during the experiment. Two more participants reported relevant issues during the experiment, but they did not believe they were related to the program's adherence to the guidelines. Instead, they believed the difficulties were due to their own circumstances. For instance, P4's poor internet connection made the access to the virtual machine unstable and impacted the the interaction with the virtual world.

On the other hand, P6 mentioned a delay in the software's response to his actions, such as zooming in and out and updating reflections and shadows when adding objects to the scene, which we believe may have affected his perceptions of the Real-time feedback guideline, although he did not specifically mention this connection.

> *"As I created various reflective elements, such as the fountain with water, there were some issues with the application. The tool would occasionally freeze and go for a while without responding; other times, it would stop responding and close, causing me to lose all of my work."*—P6

Only one participant made the connection between the technical issues and their perceptions of the guidelines. P3 had problems installing the virtual machine to access the Omniverse, which impacted his analysis of the Democratization guideline. For him, this meant that LaunchPad might not function properly on all computers, and that the instruction lacked sufficient information to assist him fix the issue. Even P1, who indicated minor difficulty with this step, stated that he "self-taught" himself how to accomplish it.

At the conclusion, the participants offered additional observations about the entire experience, from utilizing Ominiverse and reading the list of guidelines to responding to the Likert-scale questionnaire and taking part in the focus group interview. Throughout the experiment in collaborative mode, one participant missed seeing who was working with him since the tool did not display the person's name, number of coworkers in the same environment, position in the scene, or the object they were modifying at the moment. We speculate that this indirectly affected his opinion of the Sharing and collaboration guideline.

> *"When I was in collaborative mode, I did not know who was editing an object or which object was being edited; things just changed. For example, the tree's color suddenly changed, but I only knew that someone else had done it because I was also connected with them on a video call."*—P1

The participants also stated that there was little information about errors in LaunchPad and that it was difficult to determine their causes. Some of them were unable to perform simple operations such as undo (ctrl+z) but could not explain why. Before beginning the activity, the training also neglected to offer users with fundamental information about how to use the program, such as where to alter the camera speed and screen size for navigating in the scene. Such information would have increased user comfort.

3.4. The Pipeline of Using Design Guidelines for Evaluating Existing VR Authoring Tools

Figure 9 illustrates a pipeline containing a compilation of all the steps taken in this study to evaluate the intuitiveness of an existing VR authoring tool in accordance with the Design Science Research paradigm [19], whereas Figure 10 illustrates how these steps are applied as a guide for anyone who wishes to replicate the experiment using different VR authoring tools.
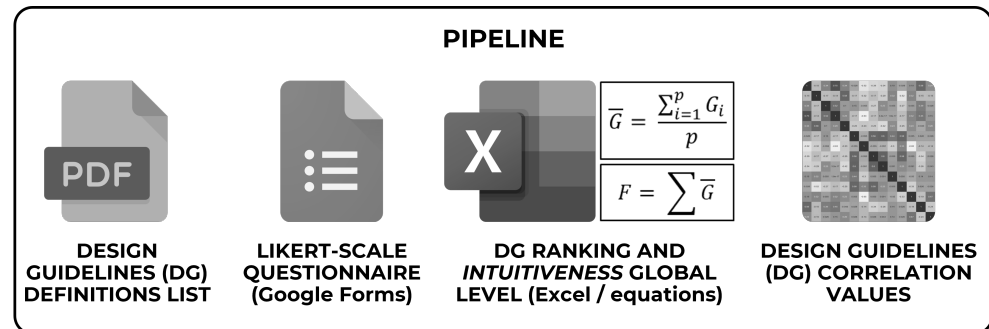

Figure 9. The pipeline and the elements that compose it (Supplementary Materials).
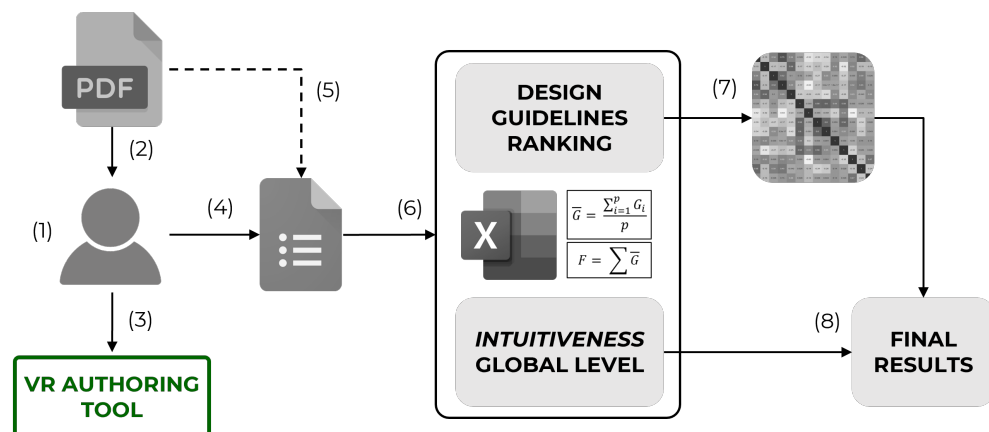
Figure 10. Process flow of the pipeline application.

Figure 10 illustrates the step-by-step process for evaluating the intuitiveness of an existing VR authoring tool using the design guidelines artifact. Different evaluators may use different-sized groups to test the to-be-evaluated tool; in the present study, six participants were utilized (1). The fourteen design guidelines definitions list should be distributed to the participants, as done here and described at the Section 2, so that they get familiar with them (2). Participants must have access to the authoring tool that will be tested and evaluated in order to complete an activity or series of tasks that demonstrate the tool's functionality (3). Hence, the Likert-Scale questionnaire can be filled independently by each participant based on their opinions of the tool's features (4). Participants must consult the design guidelines anytime they are uncertain about how to complete the questionnaire (5).

The questionnaire responses must then be analyzed so that a ranking of the scores for the design guidelines and an global level of intuitiveness may be determined. To obtain these products, he answers from the Google Forms must be exported to an Excel spreadsheet and then run through the equations in Figure 5 (6). The scores of the guidelines that form pairs of strong positive or negative correlations with others should be highlighted, as shown in Tables 2 and 3, and compared to see if the tool exhibits expected behavior (7). The final findings of the evaluation should include the ranking, the comparison with the correlation values and the intuitiveness global level, which, when combined, should reflect the intuitiveness of the evaluated VR authoring tool (8).

As the primary objective of this study was to assess the validity of the design guidelines, we utilized the focus group interview to obtain more in-depth qualitative data on them. Future experiments utilizing different VR authoring tools do not require focus group interviews into their process flow.

4. Conclusions

We demonstrated how to conduct an evaluation experiment from the perspective of the design guidelines using an existing VR authoring tool, thereby analyzing the guidelines' validity as an artifact. The proposed artifact is valuable, according to Design Science Research, because the design guidelines for virtual reality authoring tools developed by Chamusca et al. [11] perform what they are supposed to do and are operationally reliable in completing their goals. As a significant contribution to the field, we produced a pipeline encapsulating all of the steps taken in this study, which may be used as a guide for anyone desiring to recreate the experiment using the artifact in a different VR authoring tool.

The study concentrated on illustrating *how to* use the design guidelines rather than offering a wide range of quantitative data analysis. Despite the fact that the primary goal of the experiment was to qualitatively assess the validity of the design guidelines in evaluating existing VR authoring tools, the quantitative results showed that the exemplary use case still does not have a high level of intuitiveness, receiving a score of 64%, which was supported by previous feedback from users who tested the NVIDIA Omniverse Enterprise tool [30].

The correlation analysis between the guidelines sought to determine the level of interdependence between the guidelines under review, as they did not exist in isolation in any of the VR authoring tools which has the potential to be evaluated. As a result, the correlations were employed as a cross-check indicator when analyzing the findings of the Likert-scale questionnaire and focus group interviews. The cross-check confirmed that the majority of the guidelines scores behaved as predicted and that the ranking obtained using the Likert-scale questionnaire was consistent with the Omniverse functionalities.

The participants understood the definition of the guidelines and could correctly identify their existence during the experiment. The Likert-scale questionnaire provided a simple method of gathering participants' perspectives on which guidelines they agreed or disagreed about having found in the tool. Later in the focus group session, they were asked to reaffirm their viewpoint on which guidelines were easier or more difficult to identify. Comparing the responses, the easy-to-identify guidelines were connected with those that obtained the highest scores, and the difficult-to-identify guidelines with those that received low scores. This outcome was consistent with the profile of the group used in this experiment, which lacked technical capabilities and indicated that the participants' evaluation was carried out mostly using the practical examples supplied by the guidelines' definitions as direct references.

As a result, everything that the participants observed in the tool and was presented in the design guidelines definition as a practical example acted as a motivator for a rise in score, while the opposite also occurred. Therefore , when an example was not displayed in Omniverse, the definition of the guidelines became more subjective in the participants' eyes, because it could not be viewed in an illustrated and practical manner. This is supported by the participants' statement highlighting the guidelines' weakness of not offering illustrated examples with figures.

The methods and tools used presented many pros that helped us during the experiment, such as: we used a very optimized snippet of the Omniverse tool by following the LaunchPad documents, which kept the activities straightforward for the participants, not distracting them with too much information. The virtual machines facilitate the experiment, since we didn't need to acquire extra hardware resources. the Likert-scale questionnaire provided a numerical scale easy to filter and create visual charts, while the focus group interview allowed us to understand the reason behind the questionnaire answers and get feedback to improve the design guidelines. On the other hand, a good internet connection

was needed to access the Omniverse virtual machines, which affected the results of some guidelines, the Likert-scale questionnaire could not provide qualitative data to improve the design guidelines and the interview brought many subjective data to compile and turn into useful feedback, which is why these methods complement one another.

The choice of a use case that is not particularly regarded as a VR authoring tool by its developers is a limitation of this experiment, although it is crucial to account for the lack of ontologies and taxonomies in this domain. While many programs have all of the qualities of an authoring tool, such as the IVWPs, they are not frequently declared as such. Participants' inability to experiment with creating virtual worlds using VR devices also influenced their perceptions and was a limitation of this study. The participants' profile of the group used to judge the guidelines can also be viewed as a limitation, because while the participants' lack of knowledge allows for testing how well defined the guidelines are to the point of being clear to professionals who are not in the VR area, it can also lead to feedback on subjectivity in the definition of guidelines that contains more technical terms. The participants had an engineering background, which characterizes them as final users of the tool, which is very important to evaluate a software intuitiveness, but does not provide feedback on how the software development should effectively improve. Therefore, we lack feedback on how design guidelines can help software developers evaluate an existing authoring tool with the aim of modifying it for better functions.

The experiment's goal was to create a pipeline through a qualitative review of the steps performed during the experiment, rather than to provide robust quantitative data. Given the reduced sample of participants (six) and the fact we assessed only one authoring tool, the numerical data offered in the study can be viewed as a limitation. In any case, it should not be interpreted as an invalidation of the experiment, but rather as a chance for further research.

In summary, the strengths of our study include the successful validation of VR authoring tool design guidelines previously proposed, showcasing their operational reliability and effectiveness. Additionally, we created a comprehensive pipeline detailing all study steps, offering a valuable blueprint for future experiments with various VR authoring tools. The study's emphasis on the practical application of these guidelines provided a qualitative insight into their effectiveness. On the other hand, our study's limitations stem from its primary focus on qualitative evaluation, resulting in a scarcity of quantitative data analysis. This approach led to subjective evaluations of the guidelines, particularly from participants with limited technical expertise. The selection of a use case not commonly recognized as a VR authoring tool, coupled with the participants' non-technical backgrounds, introduced further subjectivity, especially in assessing more technical guidelines. Moreover, the small sample size and the examination of only one authoring tool challenge the findings' generalizability, indicating the need for expanded research.

In terms of future research, we propose altering the group of evaluators with VR industry players, such as expert programmers and designers of virtual reality experiences, as well as to carry out the experiment with more participants, to gather additional technical input. Also in terms of evaluators, important feedback can also come from people who are not part of the professional market, such as children and the elderly, or professionals who do not usually require the use of 3D creation software in their daily lives, such as teachers, doctors or lawyers. Considering that design guidelines aim to influence the development of authoring tools more intuitive for different audiences, feedback from people with very little contact with 3D creation should bring an extreme counterpoint to the feedback from designers and developers in the area.

Furthermore, we recommend experimenting with various VR authoring tools or in a context that enables the experience to be enjoyed not only on 2D screens but also on head-mounted displays. Comparing the findings of the evaluation through design guidelines with common methods for measuring usability, such as the System Usability Scale (SUS) and the After-Scenario Questionnaire (ASQ), can be used to demonstrate their efficacy as a method. The Omniverse tool can be assessed again to test if the score given for the design

guidelines is restricted to the activity outlined in the LaunchPad, as well as to examine its potential for metaverse creation and industrial applications.

In future research, the guidelines' definitions could be improved by reorganizing the list format, using pictures to explain the definitions in text, and including more explanation for the technical terms. Further tests with the design guidelines are recommended in order to propose an organization of these by different weights, resulting in different relevance among the fourteen listed today in terms of intuitiveness. Suggestions on how to search for the guidelines in an authoring tool can also be added to the document listing the guidelines (Figure 10, item 2), as participants did this in different ways in the experiment, taking notes throughout the activities or specifically looking for the examples cited for each guideline in the document. Guidelines must be implemented to direct the creation of new VR authoring tools from the outset of a software project, which may also provide feedback for their usage in the assessment of current ones. Furthermore, since immersive technologies will improve in terms of hardware and software, as well as product and service, the design guidelines definition and practical examples must reviewed and modified over time.

Through the experiment the design guidelines have proven to be beneficial even in assisting experts outside of the field on their very first interaction with tools such as Omniverse. We found that design guidelines might be useful not only in the creation of new intuitive VR authoring tools, but also in assessing the intuitiveness of current ones. As a result, the design guidelines help to make accessible tools for authoring virtual worlds that can be experienced in virtual reality, immediate effect on the creation of ontologies and the faster dissemination of technological advances like the metaverse, as a greater number of individuals from different fields of expertise become able of developing them. In terms of impact on society, the design guidelines lead to more intuitive VR authoring tools, which consequently makes them more easily disseminated among professionals from different areas who can start using VR as a resource in their day-to-day work. Areas such as education (SDG4), healthcare (SDG3), justice (SDG16) and economy (SDG8 and SDG12) are already impacted by the use of VR and have the potential to be even more impacted if specialist professionals in this context have the autonomy to use directly VR authoring tools to prototype your projects.

For a genuinely sustainable society, more individuals must be empowered to not only utilize but also create immersive virtual reality experiences, and a more sustainable virtual reality will undoubtedly help build a more sustainable future.

Omniverse tool accessible to our study, to Fabrício Vinicius de Jesus Santos for insightful comments and for Lucas Arnaldo Alves for the creation of the Correlation Graphic presented as Figure 7.

References

1. Jauhiainen, J.S.; Krohn, C.; Junnila, J. Metaverse and Sustainability: Systematic Review of Scientific Publications until 2022 and Beyond. *Sustainability* **2023**, *15*, 346. https://doi.org/10.3390/su15010346.
2. Johri, A.; Joshi, P.; Kumar, S.; Joshi, G. Metaverse for Sustainable Development in a bibliometric analysis and systematic literature review. *J. Clean. Prod.* **2024**, *435*, 140610. https://doi.org/10.1016/j.jclepro.2024.140610.
3. Cassola, F.; Pinto, M.; Mendes, D.; Morgado, L.; Coelho, A.; Paredes, H. A novel tool for immersive authoring of experiential learning in virtual reality. In Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 27 March–1 April 2021; pp. 44–49.
4. Nebeling, M.; Speicher, M. The trouble with augmented reality/virtual reality authoring tools. In Proceedings of the 2018 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), Munich, Germany, 16–20 October 2018; pp. 333–337.
5. Zhang, L.; Oney, S. Flowmatic: An immersive authoring tool for creating interactive scenes in virtual reality. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual, 20–23 October 2020; pp. 342–353.
6. Sherman, W.R.; Craig, A.B. *Understanding Virtual Reality: Interface, Application, and Design*; Morgan Kaufmann: Burlington, MA, USA, 2018.
7. Cai, Y.; See, S. *GPU Computing and Applications*; Springer: Berlin/Heidelberg, Germany, 2015.
8. Ipsita, A.; Li, H.; Duan, R.; Cao, Y.; Chidambaram, S.; Liu, M.; Ramani, K. VRFromX: From scanned reality to interactive virtual experience with human-in-the-loop. In Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021.
9. Krauß, V.; Boden, A.; Oppermann, L.; Reiners, R. Current practices, challenges, and design implications for collaborative AR/VR application development. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021.
10. Yigitbas, E.; Klauke, J.; Gottschalk, S.; Engels, G. VREUD-an end-user development tool to simplify the creation of interactive VR scenes. In Proceedings of the 2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), St. Louis, MO, USA, 10–13 October 2021; pp. 1–10.
11. Chamusca, I.L.; Ferreira, C.V.; Murari, T.B.; Apolinario, A.L.; Winkler, I. Towards Sustainable Virtual Reality: Gathering Design Guidelines for Intuitive Authoring Tools. *Sustainability* **2023**, *15*, 2924.
12. Coelho, H.; Monteiro, P.; Gonçalves, G.; Melo, M.; Bessa, M. Authoring tools for virtual reality experiences: A systematic review. *Multimed. Tools Appl.* **2022**, *81*, 28037–28060.
13. Zikas, P.; Papagiannakis, G.; Lydatakis, N.; Kateros, S.; Ntoa, S.; Adami, I.; Stephanidis, C. Immersive visual scripting based on VR software design patterns for experiential training. *Vis. Comput.* **2020**, *36*, 1965–1977.
14. Lynch, T.; Ghergulescu, I. Review of virtual labs as the emerging technologies for teaching STEM subjects. In Proceedings of the INTED2017—11th International Technology, Education and Development Conference, Valencia, Spain, 6–8 March 2017; Volume 6, pp. 6082–6091.
15. O'Connor, E.A.; Domingo, J. A practical guide, with theoretical underpinnings, for creating effective virtual reality learning environments. *J. Educ. Technol. Syst.* **2017**, *45*, 343–364.
16. Wang, Y.; Su, Z.; Zhang, N.; Xing, R.; Liu, D.; Luan, T.H.; Shen, X. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Commun. Surv. Tutorials* **2022**, *25*, 319–352.
17. Ball, M. *The Metaverse: And How It Will Revolutionize Everything*; Liveright Publishing: New York, NY, USA, 2022.
18. Pressman, R.S. *Software Engineering: A Practitioner's Approach*; Palgrave Macmillan: London, UK, 2021.
19. Gregor, S.; Hevner, A.R. Positioning and presenting design science research for maximum impact. *MIS Q.* **2013**, *37*, 337–355.
20. NVIDIA and Partners Build Out Universal Scene Description to Accelerate Industrial Metaverse and Next Wave of AI. Available online: https://nvidianews.nvidia.com/news/nvidia-and-partners-build-out-universal-scene-description-to-accelerate-industrial-metaverse-and-next-wave-of-ai (accessed on 22 February 2023).
21. Creswell, J.W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*; Sage Publications: Thousand Oaks, CA, USA, 2017.
22. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77.
23. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Syst. Rev.* **2021**, *10*, 1–11.

24. Booth, A.; Sutton, A.; Clowes, M.; Martyn-St James, M. *Systematic Approaches to a Successful Literature Review*; Sage: Thousand Oaks, CA, USA, 2021.
25. Chamusca, I.L.; Santos, F.V.D.J.; Ferreira, C.V.; Murari, T.B.; Junior, A.L.A.; Winkler, I. Evaluation of design guidelines for the development of intuitive virtual reality authoring tools: a case study with NVIDIA Omniverse. In Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, 17–21 October 2022; pp. 357–362.
26. Ning, H.; Wang, H.; Lin, Y.; Wang, W.; Dhelim, S.; Farha, F.; Ding, J.; Daneshmand, M. A Survey on Metaverse: The State-of-the-art, Technologies, Applications, and Challenges. *arXiv* **2021**, arXiv:2111.09673.
27. CES 2022: Nvidia Unleashes Omniverse as Metaverse Design Push Gains Steam. Available online: https://www.fierceelectronics.com/embedded/ces-2022-nvidia-unleashes-omniverse-metaverse-design-push-gains-steam (accessed on 17 January 2023).
28. Xu, M.; Ng, W.C.; Lim, W.Y.B.; Kang, J.; Xiong, Z.; Niyato, D.; Yang, Q.; Shen, X.S.; Miao, C. A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges. *IEEE Commun. Surv. Tutorials* **2022**, *25*, 656–700.
29. NVIDIA Omniverse: The Useful Metaverse. Available online: https://www.forbes.com/sites/karlfreund/2022/10/07/nvidia-omniverse-the-useful-metaverse/?sh=1e270893359a (accessed on 17 January 2023).
30. Building the Metaverse Using NVIDIA Omniverse. Available online: https://www.deltatre.com/insights/building-metaverse-using-nvidia-omniverse (accessed on 17 January 2023).
31. Park, S.M.; Kim, Y.G. A Metaverse: Taxonomy, components, applications, and open challenges. *IEEE Access* **2022**, *10*, 4209–4251.
32. Brooke, J. SUS: A retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.